Keystroke Pressure-Based Typing Biometrics Authentication System Using Support Vector Machines

Wahyudi Martono, Hasimah Ali, and Momoh Jimoh E. Salami

Intelligent Mechatronics System Research Group, Department of Mechatronics Engineering,
International Islamic University Malaysia, PO BOX 10, 50728, Kuala Lumpur, Malaysia
{wahyudi,momoh}@iiu.edu.my, csimah2002@yahoo.com

Abstract. Security of an information system depends to a large extent on its ability to authenticate legitimate users as well as to withstand attacks of various kinds. Confidence in its ability to provide adequate authentication is, however, waning. This is largely due to the wrongful use of passwords by many users. In this paper, the design and development of keystroke pressure-based typing biometrics for individual user's verification which based on the analysis of habitual typing of individuals is discussed. The combination of maximum pressure exerted on the keyboard and time latency between keystrokes is used as features to create typing patterns for individual users so as to recognize authentic users and to reject impostors. Support vector machines (SVMs), which is relatively new machine learning, is used as a pattern matching method. The effectiveness of the proposed system is evaluated based upon False Reject Rate (*FRR*) and False Accept Rate (*FAR*). A series of experiment shows that the proposed system is effective for biometric-based security system.

Keywords: Biometric, security, keystroke, maximum pressure, time latency and support vector machine.

1 Introduction

Almost all the activities rely on the use of computer technology. Thus, computer has become an integral part of nearly in every aspect of societal activities. The communication, aviation and financial services are already controlled by computer. People entrust with vital information such as medical and criminal records, manage transactions, pay bills and write personal letters. However, this increasing dependency on computers coupled with growing emphasis on global accessibility in cyberspace, has unveiled new threats to computer system security [1]. In addition, crimes and impostors in the cyberspace appear are almost everywhere. Crimes on the computer networks may cause serious damages, including communication blocking, perusal of classified files, commerce information destruction etc [2].

Traditional methods such as passwords and PINs are no longer adequate, as either of these can be cracked, possibly breaking to the computer system. Consequently, alternatives to traditional access control methods are in high demand. Although, a variety of authentication devices to verify a user's identity are in use, password

O. Gervasi and M. Gavrilova (Eds.): ICCSA 2007, LNCS 4706, Part II, pp. 85–93, 2007.

technique has been and will remain the preferred method. Password authentication is an inexpensive and familiar paradigm that most operating systems support. However, the confidence in ability to provide highly secured authentication is weakening. This is largely due to the wrongful use of passwords by many users and to the inhibited simplicity of the mechanism which makes it susceptible to extraordinary intruder attacks. Methods are needed, therefore, to extend and enhance the life of password techniques [3].

A software methodology that improves security by using typing biometrics has been developed to reinforce password-authentication mechanisms [3].Typing biometric or keystroke dynamics is the analysis of a user's keystroke patterns. This relies on the fact that, each user has a unique way of using the keyboard to enter a password; for example, each user types the characters that constitute the password at different speeds. In developing a scheme using keystroke dynamics for identity verification, it is very necessary to determine which keystrokes characterize the individual's key pattern. Willem et al. [3] employed fuzzy logic to measure the users typing biometric. However there are many adjustable elements such as membership functions and fuzzy rules. Although it has been claimed that many adjustable elements increase the flexibility of the fuzzy-based authentication, they also increase the complexity in designing fuzzy-based authentication system.

Taking the advantages of habitual typing of individuals possesses, this paper has proposed the design and development of keystroke-pressure based typing biometric as authentication system and to use support vector machines (SVMs) as pattern matching procedure for identifying the authorized and unauthorized user. The paper examines the use of maximum pressure exerted on the keyboard and time latency between keystrokes as features to create typing patterns for individual users. Pressure signals which are taken from underneath the keypad are extracted accordingly. Both features are then used to recognize authentic users and to reject impostors. The performance of proposed system is evaluated based on False Rejection Rate (*FRR*) and False Acceptance Rate (*FAR*).

2 Proposed System Description

Figure 1 shows the proposed Biometric Authentication System which is based on keystroke pressure-based typing biometric. The proposed system is sensitive to the pressure applied on each keystroke. It consists of the following devices:

1. Alphanumeric keyboard (Biokeyboard) embedded with force sensors to measure the pressure while typing.
2. Data Acquisition System (DAS) which consists of analog interface and DAQ hardware.
3. PC/Central Processing Unit (CPU).

This system employs special force sensors to measure the exact amount of pressure a user exerts while typing, signal processing is then carried out to construct a waveform pattern for the password entered. The maximum pressure is extracted from the waveform pattern and it is used as one of the features to authenticate the user. In addition, the proposed system also measures the actual timing traces called "latency"

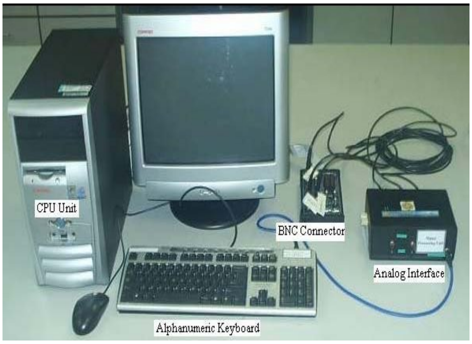


Fig. 1. Integration of biometric authentication system component

(the time between keystrokes). This proposed system in general makes four possible decisions; the authorized person is accepted, the authorized person is rejected, the unauthorized person (impostor) is accepted and the unauthorized person (impostor) is rejected. The accuracy of the proposed system is then specified based on the rate in which the system makes the decision to reject the authorized person and to accept the unauthorized person. False Rejection Rates (*FRR*) is used to measure the rate of the system to reject the authorized person whereas the False Acceptance Rates (*FAR*) is measure the ability of the system to accept the unauthorized person. Both performances are can be expressed as:

$$\mathrm{FRR} = \frac{\mathrm{NFR}}{\mathrm{NAA}} \mathrm{x}100\% \,, \tag{1}$$

$$\mathrm{FAR} = \frac{\mathrm{NFA}}{\mathrm{NIA}} \mathrm{x}100\% \,. \tag{2}$$

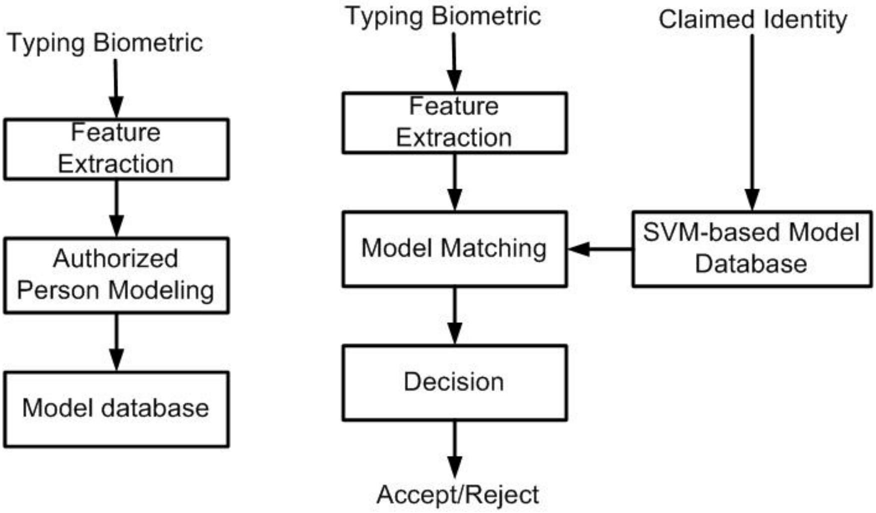
NFR is referred to the numbers of false rejections and *NFA* is referred to the number of false acceptances respectively while *NAA* and *NIA* refer to the number of the authorized person attempts and the number of impostor person attempts respectively [4]. In this paper the main objective is to develop a system that would have both low *FRR* and *FAR* as well as to achieve both high usability and high security of the system.

3 Keystroke Pressure-Based Typing Biometric System

Most applications of keystroke dynamics are in field of verification. As other methods of biometric-based security system, there are two phases in the proposed system. First phase is training or enrollment phase as shown in Fig. 2 (a) and second phase is testing as shown in Fig. 2 (b).

During the first phase, each person has to register as an authorized person by entering appropriate information in the experimental system. Each of user passwords should consist of six (6) digits. Then the passwords that contain typing biometric data are extracted. The features extracted from typing biometric of the password are used to develop models of the authorized persons.

The second phase in the proposed system is testing or operational phase as shown in Fig. 2(b). In the second phase, an attempt would be made to access the system when a user would be required to enter his/her password. At the same time, the system computes the person's typing for the password just entered. It then compares this with the claimed person model to verify his/her claim.



(a) Enrolment phase (b) Operational phase

Fig. 2. Basic structure of keystroke pressure-based biometric authentication system

In this phase also, there is decision process in which the system decides whether the feature extracted from the given typing pattern of password matches with the model of the claimed person. In order to give a definite answer of access acceptance or rejection, a threshold is set. When degree of similarity between a given passwords is greater than threshold, the system grants access to the person, otherwise the system will reject access to the system.

3.1 Feature Extraction

Feature extraction is the process whereby unique data are extracted from the sample and a template is created. The templates for any two persons should differ whereas different samples for the same person should be identical [4]. As shown in Fig. 2, feature extraction is one of the important processes in the proposed system. Feature extraction is the process of converting the biometric data to feature vector which can be used for classification.

There are several different features of the keystroke dynamics which can be used when the user presses the keyboard keys. Possible features include [5]:

1. Latency between consecutive keystrokes.
2. Duration of the keystroke, hold-time.
3. Overall typing speed.
4. Frequency of errors (how often the user has to use backspace).
5. The habit of using additional keys in the keyboard, for example writing numbers with the numpad.
6. The order that user press keys when writing capital letters, (is shift or the letter key released first?).
7. The force used when hitting keys while typing (requires a special keyboard).

In the proposed system, combine features of maximum pressure and latency are adopted as the features since this features combination is considerably the effective feature to be used in the keystroke-based authentication system [6]. The alphanumeric keyboard with additional press sensor, measures the person's biometric data during the process of identifying oneself.

4 Support Vector Machines

Support vector machine (SVMs) is a relatively new learning machine technique, which is based on the principle of structural risk minimization. A SVM is binary classifier that optimally separates the two classes. There are two important aspects in the development of SVM as classifier. The first aspect is determination of the optimal hyperplane which will optimally separate the two classes and the other aspect is transformation of non-linearly separable classification problem into linearly separable problem. This section will discuss in brief the two aspects of the SVM development. Detail discussion on the SVM can be found in the introductory text by Burges [7], for more detail description, Cristianini and Shawe [8].

Fig. 3 shows linearly separable binary classification problem with no possibility of miss-classification data. Let $\boldsymbol{x}$ and y be a set of input feature vector and the class label repectively. The pair of input feature vectors and the class label can be represented as tuples $\{x_i,y_i\}$ where $i=1,2,\cdots,N$ and $y=\pm 1$. In the case of linear separable problem, there exists a separating hyperplane which defines the boundary between class 1 (labeled as y = 1) and class 2 (labeled as y = -1). The separating hyperplane is;

$$w \cdot x + b = 0 \quad , \tag{3}$$

which implies

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1,2,\cdots,N \quad . \tag{4}$$

Basically, there are numerous possible values of $\{\mathbf{w},b\}$ that create separating hyperplane. In SVM only hyperplane that maximizes the margin between two sets is used. Margin is the distance between the closest data to the hyperlane. Referring to Fig. 4, the margins are defined as d_+ and d_-. The margin will be maximized in the case

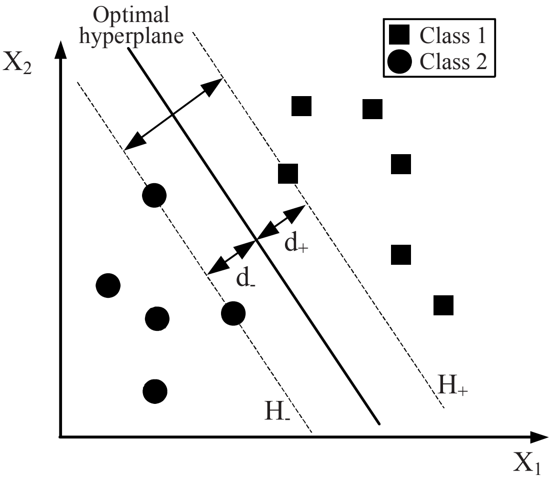


Fig. 3. SVM with linear separable data

$d_+ = d_-$. Moreover, training data in the margins will lie on the hyperplanes H_+ and H_-. The distance between hyperplane H_+ and H- is

$$d_+ + d_- = \frac{2}{\|w\|} \quad (5)$$

As H_+ and H_- are the hyperplane in which the closest training data to the optimal hyperplane, then there is no training data which fall between H+ and H-. This means the hyperplane that separates optimally the training data is the hyperplane which minimizes $\|w\|^2$ so that the distance of Eq. (5) is maximized. However, the minimization of $\|w\|^2$ is constrained by Eq. (4). When the data is non-separable, slack variables,ξ_i, are introduced into the inequalities for relaxing them slightly so that some points allow to lie within the margin or even be misclassified completely. The resulting problem is then to minimize

$$\frac{1}{2}\|w\|^2 + C\left(\sum_i L(\xi i)\right) \quad (6)$$

Where C is the adjustable penalty term and L is the loss function. The most common used loss function is linear loss function, $L(\xi_i) = \xi_i$.

The optimization of Eq. (13) with linear loss function using Lagrange multipliers approach is to maximize:

$$L_D(\mathbf{w}, b, \alpha) = \sum_i^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle, \quad (7)$$

subject to

$$0 \le \alpha_i \le C \quad , \tag{8.a}$$

$$\sum_i^N \alpha_i y_i = 0 \tag{8.b}$$

In which α_i is the Lagrange multipliers. This optimization problem can be solved by using standard quadratic programming technique. Once the problem is optimized, the parameters of optimal hyperplane are

$$\mathbf{w} = \sum_i^N \alpha_i y_i \mathbf{x}_i \quad , \tag{9}$$

As matter of fact, α_i is zero for every $\mathbf{x}_i$ except the ones that lie on the margin. The training data with non-zero α_i are called as support vectors. In the case of a non-linear separable problem, a kernel function is adopted to transform the feature space into higher dimensional feature space in which the problem become linearly separable. Typical kernel functions commonly used are listed in Table 1.

Table 1. Kernel function commonly used in SVM

Kernel	$K(x,x_j)$
Linear	$x^{\mathbf{T}} \cdot x_j$
Polynomial	$(x^{T} \cdot x_j + 1)^d$
Gaussian RBF	$\exp(-\|x-x_j\|2/2\sigma^2)$

5 Results

5.1 Experimental Setup

To evaluate the effectiveness of the proposed keystroke pressure-based typing biometric authentication system, a group of five (5) persons is used as data collection in the experiment. Three (3) persons are considered as authorized person and the other two (2) persons are assumed as imposters. Each person has been required to type six characters of their own password for 200 times, 100 sets are used for training whereas the rest are used for testing data. So, each typed character of the password has maximum pressure and latency (time between keystroke). For example, when the password "asd123" were entered, then the time duration between the letter pairs (a-s), (s-d), (d-1), (1-2) and (2-3) would be computed. In addition, each character has its own maximum pressure. Here, each typed character of the password consists of five latency and six maximum force features, resulting in eleven features for each user.

The performance of biometric systems is usually described by two error rates: (*FRR*) and (*FAR*). Hence, the effectiveness, of the proposed system in testing (operational) phase is evaluated based upon *FRR* and *FAR*. The *FAR* is calculated based on the close set and open set. In the close set, the authorized person uses other authorized person identity and password to access the system. On other hand, the open set is referred impostor accesess.

5.2 Training of SVM-Based Models

A SVM with polynomial kernel function of order 5 is used for developing person models. Each authorized person has its own SVM-based model characterized by a set of support vectors. The support vectors are obtained by using quadratic programming in the MATLAB environment. The penalty term C of 100 is used to anticipate misclassified data. Table 2 shows the training time and the classification rate when the SVM is used to develop all authorized person model based on their typing biometric.

Table 2. Training performance

Authorized Person	Training Time (sec)	Classification Rate (%)
Person 1	0.0625	100
Person 2	0.6563	100
Person 3	0.4687	100
Person 4	0.6250	100
Person 5	0.2815	100
Average	0.4188	100

As shown in Table 1, all of the SVM-based speaker models give perfect classification rates which are 100%. There are no errors in identifying the authorized persons based on the typing biometric data used in training phase. Furthermore, all of the SVM-based model of authorized persons can be trained in a very short time which is less than 0.5 second in average. Therefore, it can be concluded that the SVM models are promising to be used in the proposed access control system.

5.3 Testing of SVM-Based Models

Further experiment is carried out using testing data which are not used during modeling (training) process. Table 3 shows the performances of the SVM-based authorized models examined by using testing data. The SVM-based authentication results in a good average *FRR* which is 5 % in average. The maximum *FRR* is 19%, which is quite high, is for Person 3. Therefore, the proposed system produces a low *FRR* for the entire authorized person, except Person 3. A low *FRR* means the SVM-based authentication is conveniently used by the authorized persons. Further *FRR* improvement has to be done, especially to improve *FRR* of Person 3.

Furthermore, the system also produces a good *FAR* for close set condition. The *FAR* of the close set is very good, which is less than 1 % in average. The maximum *FAR* for close set is 1.75%. This means the authorized person is difficult to use other authorized person identity to access to door. However, there is a problem with *FAR* of open set. Although the average *FAR* is quit low, which is 14.7%, the *FAR* of Person 1,2 and 5, which are larger than 10%, are not acceptable. This means impostors can easily be used identity of Person 5 to access to door. Hence further improvement also has to be done for increasing the security level from the impostor by improving the *FAR*, especially for Person 1, 2 and 5.

Table 3. Testing performance

Authorized Person	FRR (%)	FAR (%)	
		Close Set	Open Set
Person 1	2	0	10
Person 2	7	1.75	14
Person 3	19	0.5	1.5
Person 4	0	0.75	0
Person 5	0	1.75	48
Average	5.6	0.95	14.7

6 Conclusions

This paper has examined development of keystroke pressure-based biometrics authentication system for security. The combining features of maximum pressure with latency are used as features to verify the authorized person due to unique typing biometric of each individual. Support vector machine (SVM) is adopted to build the authorized person model based and her/his unique typing biometric. A series of experiment shows that the proposed system that uses combined features of maximum pressure with latency is effective for biometric-based security system since it gives better false acceptance rate (*FAR*) of the closet condition and good false rejection rate (*FRR*). However, further study has to be done to improve the *FAR* for open set condition as well as increase the level of security of the system.

References

1. Eltahir, W.E., Lai, W.K., Salami, M.J.E., Ismail, A.F.: Design of a Pressure Based Typing Biometric Authentication System. In: Proceeding of the 8th Australian and New Zealand Intelligent Information System Conference ANZIIS, Sydney AU (2003)
2. Lin, D.T.: Computer-Access Authentication with Neural Network Based Keystroke Identity Verification. In: Proceeding International Conference on Neural Networks, Houston, Texas, USA, pp. 174–178 (1997)
3. William, G., De Jan, H.P.: Enhanced Password Authentication through Fuzzy Logic. IEEE Expert 12(6), 38–45 (1997)
4. Zhang, D.: Automated Biometric Technologies and System. Kluwer Academic Publishers, Dordrecht (2000)
5. Ilonen, J.: Keystroke Dynamics. In: Advanced Topics in Information Processing – Lecture (2003)
6. Ali, H., Wahyudi, Salami, M.J.E.: Keystroke Pressure-Based Typing Biometrics Authentication System Using Artificial Neural Network. In: Proceeding 1st International Conference on Control, Instrumentation and Mechatronics Engineering, Johor Bahru, Malaysia, pp. 407–412 (2007)
7. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
8. Cristianini, N., Shawe, T.J.: An introduction to Support Vector Machine and other kernel-based learning methods. Cambridge University Press, Cambridge (2000)